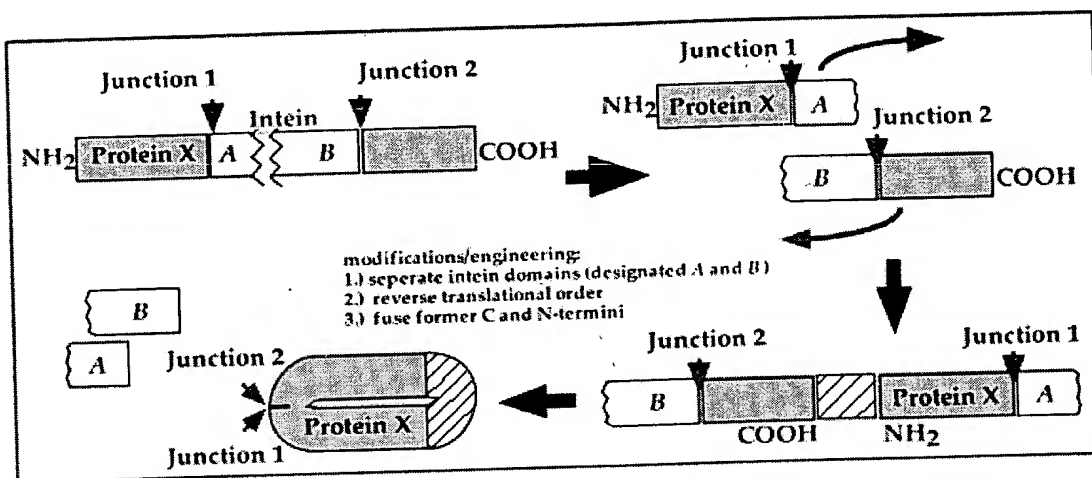
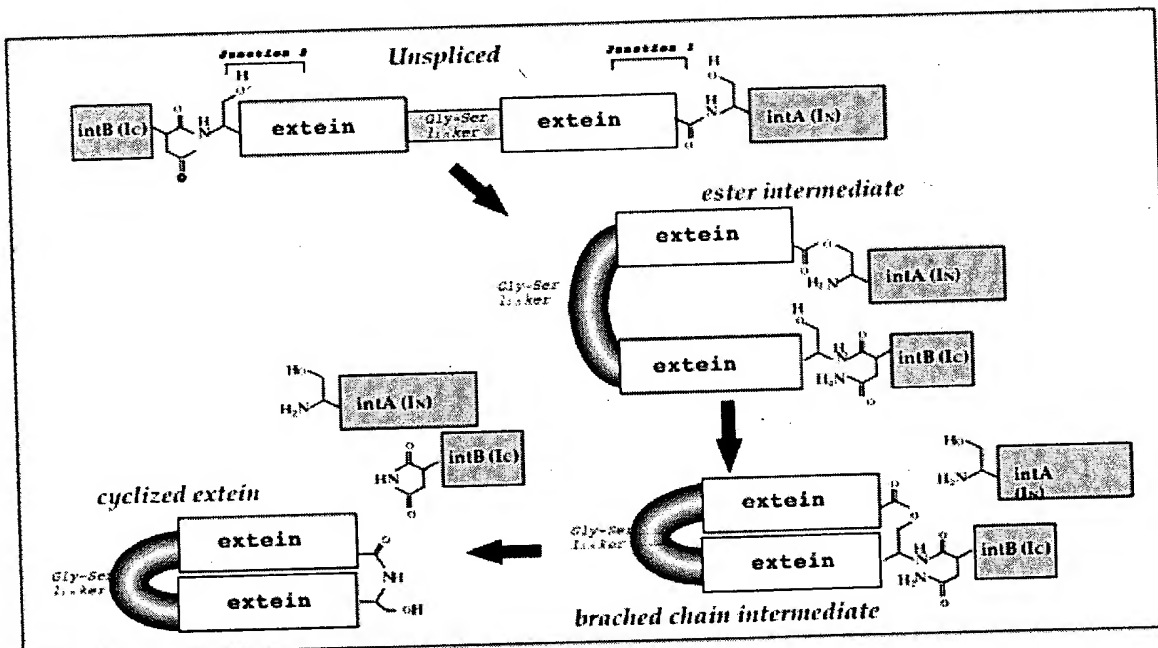


Figure 1

A



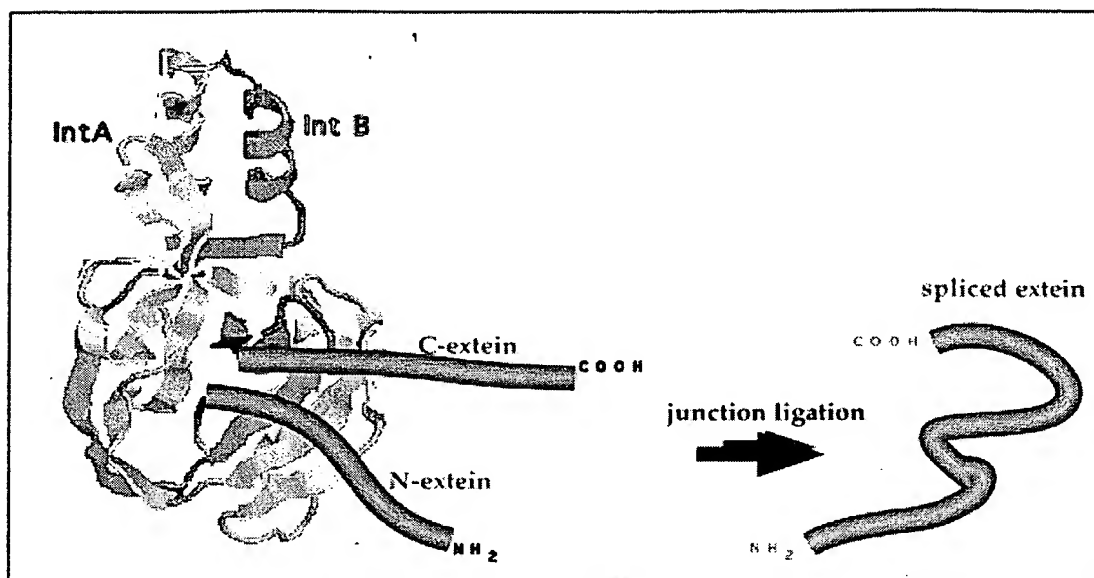
B



09400770-030601

Figure 2

A



B

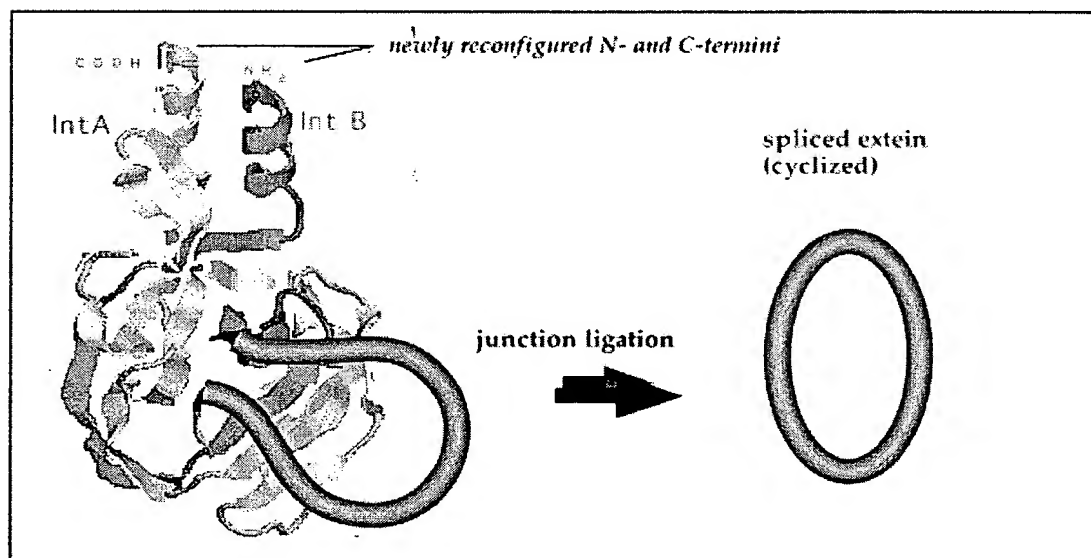


FIGURE 3

A)

GCISGDSLISLASTGKRVS IKDLLDEKDFEIWAIN EQTMKLESAKVS RVFCTGKKLVYI  
LKTRLGRTIKA  
TANHRFLTIDGWKRLDELSLKEHIALPRKLESSSLQLMSDEELG LLGHLIGDGCTLPR  
HAIQYTSNKIEL  
AEKVVELAKAVFGDQINPRISQERQWYQVYIPASYRLTHNKKNPITKWLENLDVFG  
LSYEKFVPNQVFE  
QPQRAIAIFLRHLWSTDGCVKLIVEKSSRPVAYYATSSEKLA KDVSLLKLGINARL  
SKISQNGKGRDN  
YHVTITGQADLQIFVDQIGAVDKDKQASVEEIKTHIAQHQA NTNRDVIPKQIWKTYV  
LPQIQIKGITTRD  
LQMRLGNAYCGTALYKHNLSRERAAKIA TITQSPEIEKLSQSDIYWDSIVSITETGVEE  
VFDLTVPGPHN  
FVANDIIVHNS

B)

YCITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHG NPNVLADRLFHSGEHP  
VYTVRTVEGLRV  
TGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAP  
TTYTVGVPGLVRF  
LEAHRDPDAQAIADELTDGRFYAKVASVTDAGVQPVYSLRVDTADHAFITNGFV  
SHNT

C)

ECLTSDHTVL TTRGWIPIADVTLDDKVAVL DNNTGEMSYQNPQKVHKYDYEGPMY  
EVKTAGVDL FVTPNH  
RMYVNTTNNTTNQNYNLVEASSIFGKKVRYKND AIWNKTDYQFILPETATLTGHTN  
KISSTPAIQPEMNA  
WLTF FGLWIANGHTTKIAEKTAENNQQKQRYKVILTQVKEDVCDIIEQTLNKLGFNFI  
RSGKDYTIENKQ  
LWSYLNPF DNGALNKYLPDWVWELSSQQCKILLNSLCLGNCLFTKNDDTLHYFSTS  
ERFANDVSRLALHA  
GTTSTIQLEAAPS NLYDTIIGLPVEVNTTLWRVIINQSSFY SYSTDKSSALNLSNNVAC  
YVNAQSALTLE  
QNSQKINKNTLV LTKNNVKSQTMHSQRAERVD TALLTQKELDNSLNHEILINKNPGT  
SQLECVVNPEVNN  
TSTNDRFVYYKGPVYCLTGPNNVFYVQRNGKAVWTGNS

FIGURE 3

D)

LCVAPETMILTEDGQFPIKDLEGKIIKVWNGNEFSSVTVVKTGTEKELLELEVELSNGCT  
LSCTPEHKFIIV  
KSYTEAKKQKTDDNAIANAERVDAQDLKPRMKLIKFDLPTLFGNSEHDIKYPYTHGF  
FCGDGTYTKYGKP  
QLSLYGDKKELLTYLDVRTMTGLEDA SGRLNTWLPLDLAPKFDVPINSSLECRM EW  
LAGYLDADGCVFRN  
GTNESIQVSCIHLDFLKRIQLLLIGMGVTSKITKLHDEKITTMPDGKGGQKPYSCKPIW  
RLFISSSGLYH  
LSEQGFETRRLKWEPRQPQRNAERFVEVLKVNKTGRVDDTYCFTEPINHAGVFNGIL  
TGQC

E)

GCFTKGTQVMMADGADKSIESIEVGDKVMGKDGM PREVVGLPRGYDDMYKVRQL  
SSTRNAKSEGLMDFT  
VSADHKLILKTKQDVKIA TRKIGGNTYTGVTFYVLEKTKTGIELVKAKTKVFGHHIH  
GQNGAEKKAATFA  
AGIDSKEYIDWIEARDYVQVDEIVKTSTTQMINPVHFESGKLGNWLHEHKQNKSLA  
PQLGYLLGTWAGI  
GNVKSSAFTMNSKDDVKLATRIMNYSSKLGMTCSSTESGELNVAENEEFFNNLGA  
EKDEAGDFTFDEFT  
DAMDELTINVHGAAASKKNLLWNALKSLGFRAKSTDIVKSIPQHIAVDDIVVRESLI  
AGLVDAAGNVET  
KSNGSIEAVVRTSFRHVARGLVKIAHSLGISSINIKDTHIDAAGVRQEFACIVNLTGA  
PLAGVLSKCAL  
ARNQTPVVKFTRDPVLFNFDLIKSAKENYYGITLAEETDHQFLLSNMALVHNC

F)

GCLSYATNQPYFLKSDNVNFSKLTSLKVS NHYILSATLELLIPFQYNRIYPIVSLIKREL  
QTGYKV VYEL  
DFYISVIVSTVEHYVLT LN GWKRILELTVDDL VATLDIQYLIYNNTEVDLFSSNVIFSS  
VINLICMNRIN  
VYDFWIPKTNNFFVNALLVHNS

G)

GCISKFSHIMWSHVSKPLFNFSIKKSHMHNFNKNIYQLLDQGEAFISRQDKKTTYKIR  
TNSEKYLELTSN  
HKILTLRGWQRCDQLLCNDMITTQIGFELSRKKKYLLNCIPFSLCNFETLANINISNFQ  
NVFDFAANPIP  
NFIANNIIVHNS

FIGURE 3

H)

GCFAGTNNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQ  
HRAHKSDSSREVPE  
LLKFTCNATHEL VVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVS  
KSYPISGPERA  
NELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMOK  
SKFHLTIEGPKVLA  
YLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAYKDRKEPQVAKTVN  
LYSKVVRGNGIRNN  
LNTENPLWDAIVGLGFLKDGVKNIPLSTDNIGTRETFLAGLIDSDGYVTDEHGIKA  
TIKTIHTSVRDG  
LVSLARSLGLVVSNAEPAKVDMNGTKHKISYAIYMSGGDVLLNVLSKCAGSKKFR  
PAPAAFARECRGF  
YFELQELKEDDYGITLSDDSDHQFLLANQVVVHNC

I)

GCFAYGTRGALADGTTEKIGKIVNQKMDVEVMSYDPDQVVPKVVNWFNNGPA  
EQFLQFTVEKSGGNG  
KSQFAATPNHLIRTPAGWTEAGDLVAGDRVMAAEPHRLSDQQFQVVLGSLMGDGN  
LSPNRRDRNGVRFRM  
GHGAKQVDYLQWKTALLGNIKHSTHVNDKGATFVDFTPLPELAELQRAVYLGDK  
KFLSEENFKALTPLA  
LVFWYMDGPFVTRSKGLQERTAGGSGRIEIVCEAMSEGNRIRLRDYLDRDTHGLDV  
RLRLSGAAGKSVLV  
FSTASSAKFQELVAPYITPSMEYKLLPRFRGQGAVTPQFVEPTQRLVPARVLDVHVK  
PHTRSMNRFDIEV  
EGNHNYFVDGVMVHNS

J)

YCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEVLE  
YELEDGSVIR  
ATSDHRFLTDDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLDAGTIK

K)

KALALDTPLPTPTGWTAMGDVAVGDELLAVDEAPTRVVAATEVMLGRPCYEIEFSD  
GTVIVADAQHQPWT  
SYGIRTSACLRCGLDIIAAAGSTPRHAGRLTTAAFMAPVLCIDSVRRVRSVPVRCVEV  
DNAAHLYLARG  
MVPHTNS

bioRxiv preprint doi: <https://doi.org/10.1101/022089>; this version posted October 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE 3

L)

GALAYDEPIYLSDGNIINIGEFVDKFFKKYKNSIKKEDNGFGWIDIGNENIYIKSFNKLS  
LIIEDKRILR  
VWRKKYSGKLIKITTKNRREITLTHDHPVYISKTGEVLEINAEMVKVGDYIYIPKNNTI  
NLDEVIKVETV  
DYNNGHIYDLTVEDNHTYIAGKNEGFAVSNC

M)

GALYDFSVIQLSNGRFVLIGDLVEELFKKYAEKIKTYKDLEYIELNEEDRFEVVSVP  
DIKANKHVVS  
RVWRRKVREGEKLIRIKTRTGNEIILTRNHPLFAFSNGDVVRKEAEKLVGDRVAVM  
MRPPSPQTKA  
VVDPAIYVKISDYLLVPNGKGMIVPNDGIPPEKAQYLLSVNSYPVKLVREVDEKLS  
YLAGVILGDGY  
ISSNGYYISATFDDEAYMDAFVSVVSDFIPNYVPSIRKNGDYTIVTVGSKIFAEMLSRI  
FGIPRGRKS  
MWDIPDVVLSNDDLMRYFIAGLFDADGYVDENGPSIVLVTKSETVARKIWYVLQRIG  
IISTVSRVKSR  
GFKEGELFRVIISGVEDLAKFAKFIPLRHSRKRAKLMEILRTKKPYRGRRTYRVPISSD  
MIAPLRQML  
GLTVAELSKLASYYAGEKVSESLIRHIEKGRVKEIRRSTLKGIALALQQIAKDVGNEE  
AWVRAKRLQI  
IAEGDVYWDEVVSVEEVDPKELGIEYVYDLTVEDDHNYVANGILVSNC

N)

PCVSGDTIVMTSGGPRTVAELEGKPFTALIRGSGYPCPSGFFRTCERDVYDLRTREGH  
CLRLTHDHRVL  
VMDGGLEWRAAGELERGDRLVMDDAAGEFPALATFRGLRGAGRQDVYDATVYGA  
SAFTANGFIVHNC

O)

GCIDGKAKIIFENEGEEHLTTMEEMYERYKHLGEFYDEEYNRWGIDVSNVPIYVKSF  
DPESKRVVKGKVN  
VIWKYELGKDVTKEYEITNKGTKILTSPWHPFFVLTPDFKIVEKRADELKEGDILIGGM  
PDGEDYKFIFD  
YWLAGFIAGDGCDFKYHSHVKGHEYIYDRLRIYDYRIETFEIINDYLEKTFGRKYSIQ  
KDRNIYYIDIKA  
RNITSHYKLKLEGIDNGIPPQILKEGKNAVLSFIAGLFDAEGHVSNNKPGIELGMVKNRL  
IEDVTHYLNAL  
GIKARIREKLRKDGIDYVLHVEEYSSLLRFYELIGKNLQNEEKREKLEKVLSNHHKGGN  
FGLPLNFNAFKE  
WASEYGVFEKTNQSQTIAIINDERISLGQWHTRNRVSKAVLVKMLRKLYEATKDEEV

FIGURE 3

P)

NSILPEEWVPLIKNGKVKIFRIGDFVDGLMKANQGKVKKTGDTEVLEVAGIHAFSFD  
 RKSCKARVMAVKA  
 VIRHRYSGNVYRIVLNSGRKITITEGHSLFVYRNGDLVEATGEDVKIGDLLAVPRSVN  
 LPEKRERLNIVE  
 LLLNLSPEETEDIILTIPVKGRKNFFKGMLRTLRLWIFGEEKRVRTASRYLRHLENLGYI  
 RLRKIGYDIID  
 KEGLEKYRTLIEKLVDVVRYNGNKREYLVEFNAVRDVISLMPEEELKEWRIGTRNG  
 FRMGTFVDIDEDFA  
 KLLGYYVSEGSARKWKNQTGGWSYTVRLYNENDEVLDDEHMLAKKFFGKVVRGK  
 NYVEIPKKMAYIIFES  
 LCGTLAENKRVPEVIFTSSKGVRWAFLEGYFIGDGDVHPSKRVRLSTKSELLVNGLV  
 LLLNSLGVSAIKL  
 GYDSGVYRVYVNEELKFTEYRKKKNVYHSHIVPKDILKETFGKVVFQKNISYKKFREL  
 VENGKLDREKAKR  
 IEWLLNGDIVLDRVVEIKREYYDGYVYDLSVDEDENFLAGFGFLYAHNS

Q)

DSVTGETEIIKRNGKVEFVAIEELFQRVDYRIGEKEYCVLEGVEALTLDNRGRLVWK  
 SV  
 PYVMRHRTNKRIYRVWFTNSWYLDVTEHSLIGYMNTSKVKPGKPLKERLVEVKPG  
 ELGE  
 SVKSLITPNRAIAHGIRVNPIAVKLWELIGLLVGDGNWGGQSNWAKYNVGLSLGLDK  
 EEI  
 EEKILKPLKNTGIISNYYDKSKKGDVSILSKWLARFMVRYFKDESGSKRIPEFMFNLP  
 RE  
 YIEAFLRGLFSADGTVSLRKGVPVRLTSVNPELSSSVRKLLWLVGVSNSMFMVETNP  
 NRY  
 LGKESGTHSVHVRKDKHRFAERIGFLLDRKATKLSNLGGHTSKKRAYKYDFDLVY  
 PKK  
 VEEIAYDGYVYDIEVEGTHRFFANGILVHNT

R)

KCLLPEEKVVLPEIGLVTLRELFELANEVVVKDEEKEVRKLGKMLTGVDERGNVKL  
 LNALYVWRVAHK  
 GEMIRVKVNGWYSVTVTPEHPFLTNRGWVKAGELKEGDYIAIPRRVYGNEDIMKFS  
 KIAKELGIKGE  
 KEFYLAGASIDIPIKVLFLAPSKLVSAFLRGYFDAKGVVRENYIEVPLFEDLPLLILRFG  
 IVSRIEKS  
 TLKISGKRNLLEFRKHVGFTDSEKAKALDELISKAKESERYPIIEELRRLGLLFGFTRN

ELRIEENPT  
YEVIMEILERIERGSPNLAEKIAVLEGGRIKEENYLRILEEGLIENGKLTGKELLEW  
RNREFDSK  
DVDYVRNIVENLVFLPVEKVERIEYEGYVYDVTTHNFVANGILVHNT

FIGURE 3

S)

QCFSGEEVIVEKGKDRKVVKLREFVEDALKEPSGEGMDGDIKVTYKDLRGEDVRIL  
TKDGFVKLLYVVK  
REGKQKLKIVNLDKDYWLAVTPDHKVFTSEGLKEAGEITEKDEIIRVPLVILDGPKI  
ASTYGEDGKFDD  
YIRWKKYYEKTGNGYKRAAKELNIKESTLRWWTQGAKPNSLKMIEELEKLNLLPLT  
SEDSRLEKVAIILG  
ALFSDGNIDRNFNLTLSFISSEKAIERFVETLKELFGEFNYEIRDNHESLGKSILFRTWD  
RRIIRFFVAL  
GAPVGNKTKVKLELPWWIKLKPSLFLAFMDGLYSGDGSVPRFARYEEGIKFNGTFEI  
AQLTDDVEKKLPF  
FEEIAWYLSFFGIKAKVRVDKTGDYKVKRLIFSQSIDNVNLFLEFIPLSPAKREKFLR  
EVESYLAAPV  
ESSLAGRIEELREHFNRIKKGERRSFIETWEVVNVTYNVTTETGNLLANGLFVKNS

T)

LCLTPDTYVVLGDGRIETIEDIVNAKERNVLSLDLDNLSIKIDTAIKFWKLRYNLGNLSK  
ITLSNNYELKA  
TPDHCLLVLRDNQLKWIPAKDIKENDYIAMPFNYKVERKPISLLNLLKYLDITDVLIE  
FDENSTIFEKIA  
EYIRNNIKTSTKYKYLRNRRVPLKYLIEWNFDLDEIEKEAKYIYKSVAGTKKIPLFKL  
DERFWYFAGLVL  
GDGSIQDSKIRIAQTPLKDVKSILDETFFPLHNWISGNQVIISNPIIAEILEKLGMRNGKL  
NGIIFSLPE  
SYINALIAGYFDTDGCFSLLYDKKAKKHNLRMVLTSKRRDVLEKIGIYLNLSIGILNTL  
HKSREVYSLIIS  
NKSLETFFEKIAKYLKIRKEAFINGYKTYKKEHEERFECDLLPVKEVFKKLTFEKGRK  
EILKDSKIHEN  
WYKEKTNIPREKLKTVLRYANNSEHKEFLEKIVNGDISFVRVKKVENIPYDGYVYD  
LSIKHNQNFISNG  
VISHNC

U)

KCLTGDTKVIANGQLFELRELVEKISGGKFGPTPVKGLKVIGIDEDGKLREFEVQYVY  
KDKTERLIRIT  
RLGRELKVTPYHPLLNNRRNGEIKWVKAELKPGDKLAVPRFLPIVTGEDPLAEWLG  
YFLGGGYADSKEN  
LIMFTNEDPLLQRFMELTEKLFSDARIREITHENGTSKVYVNSKKALKLVNSLGNAH



IPKECWGRGIRSF  
 LRAYFDCNGGVKGNAIVLATASKEMSQEIAYALAGFGIISRIQEYRVIISGSDNVKKFL  
 NEIGFINRNKL  
 EKALKLVKKDDPGHDGLEINYELISYVKDRLRLSFFNDKRSWSYREAKEISWELMKE  
 IYYRLDELEKLKE  
 SLRGILIDWNEVAKRIEEVAEETGIRADELLEYIEGKRKLSFKDYIKIAKVLGIDVEHT  
 IEAMRVFARK  
 YSSYAEIGRRLGTWNSSVKTILESNAVNVAILERIRKIELELIEEILSDEKLKEGIAYLIF

### FIGURE 3

U) cont.

LSQNELYWD  
 EITKVEELRGEFIIYDLHVPGYHNFIAGNMPTVVHNT

V)

SCVTGDTKVYTPDEREVKIRDFMNYFENGLIKEVSNRIGRDTVIAAVSFNSRIVGHPV  
 YRLTLESGRIIE  
 ATGDHMF LTPEGWKQTYDIKEGSEVLVKPTLEGTPYEPDPRVIIDIKEFYNFLEKIERE  
 HNLKPLKEAKT  
 FRELITKDKEKILRRALELRAEIENGLTKREAEILELISADTWIPRAELEKKARISRTL  
 NQILQRLEKK  
 GYIERRIEGRKQFVRKIRNGKILRNAMDIKRILEEEFGIKISYTTVKKLLSGNVDMAY  
 RILKEVKEKWL  
 VRYDDEKAGILARVVGFILGDGHLARNGRIFWNSKEELEMLANDLRKLGLKPSEIIE  
 RDSSEIQGRKV  
 KGRIYMLYVDNAAFHALLRFWKVEVGNKTKKGYTVPEWIKKGNLFVKREFLRGLF  
 GADGTKPCGKRYNFN  
 GIKLEIRAKKESLERTVEFLNDVADLLREFDVDSKITVSPTKEGFIIRLIVTPNDANYLN  
 FLTRVGYAYA  
 KDTYARLVGEYIRIKLAYKNIILPGIAEKAIELATVTNSTYAAKVLGVSRDFVVRNLK  
 GTQIGITRDFMT  
 FEEFMKERVNLNGYVIEKVIKKEKLG YLDVYDVT CARDHSFISNGLVSHNC

W)

NCLTSNSKILTDDGYIYKLEKLKEKLDLHIKIYNTEEKSSNILFVSERYADEKIIRIK  
 TESGRVLEGS  
 KDHPVLTNGYVPMGMLKEGDDVIVYPYEGVEYEEPSDEILDEDDFAEYDKQIIKY  
 LKDRGLLPLRMDN  
 KNIGIARLLGFAGDGSIVKENGDRERLYVAFYGKRETLIKIREDLEKLGKASRIYSR  
 KREVEIRNAY  
 GDEYTSLCDNSIKITSKAFALFMHKLGMPIGKKTEQIYKIPEWIKKAPKWVKRNFLA  
 GLFGADGSRAVF  
 KNYTLPINLTMSKSEELKENILEFLNEIKLLLA EFDIESMIYEIKSLDGRVSYRLAIVG  
 EESIKNFLGR  
 INYEYSGEKKVIGLLAYEYLRRKDIAKEIRKKCIKRAKELYKKGVTVSEMLKMDEFR  
 NEFISKRLIERAV

FIGURE 3: 04.00360

X)  
KCVDGDTLVLTKEFGLIKIKELYEKLDGKGRKIVEGNEEWTELEKPITVYGYKDGKI  
VEI  
KATHVYKGVSSGMVEIRTRTGRKIKVTPIHRLFTGRVTKDGLILKEVMAMHVKPGD  
RIAV  
VKKIDGGEYIKLDSSNVGEIKVPEILNEELAEFLGYLMANGTLKSGIIEIYCDDSELLER  
VNSLSLKLFGVGGRIVQKVDGKALVIQSKPLVDVLRRLGVPEDKKVENWKVPRELL  
LSPS.

### FIGURE 3

**X) cont.**  
 NVVRAFVNAYIKGKEEVEITLASEEGAYELSYLFAKLGIVTISKSGEYYKVRVSRRG  
 NL  
 DTIPVEVNGMPKVLPLYEDFRKFAKSIGLEEVAENHLQHIIFDEVIDVRYIPEPQEVYDV  
 T  
 TETHNFVGGNMPTLLHNT

1043968

# Figure 4A

## Intein B

MESG[S]PEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVAND

*cyclic insert (with flagg epitope)*

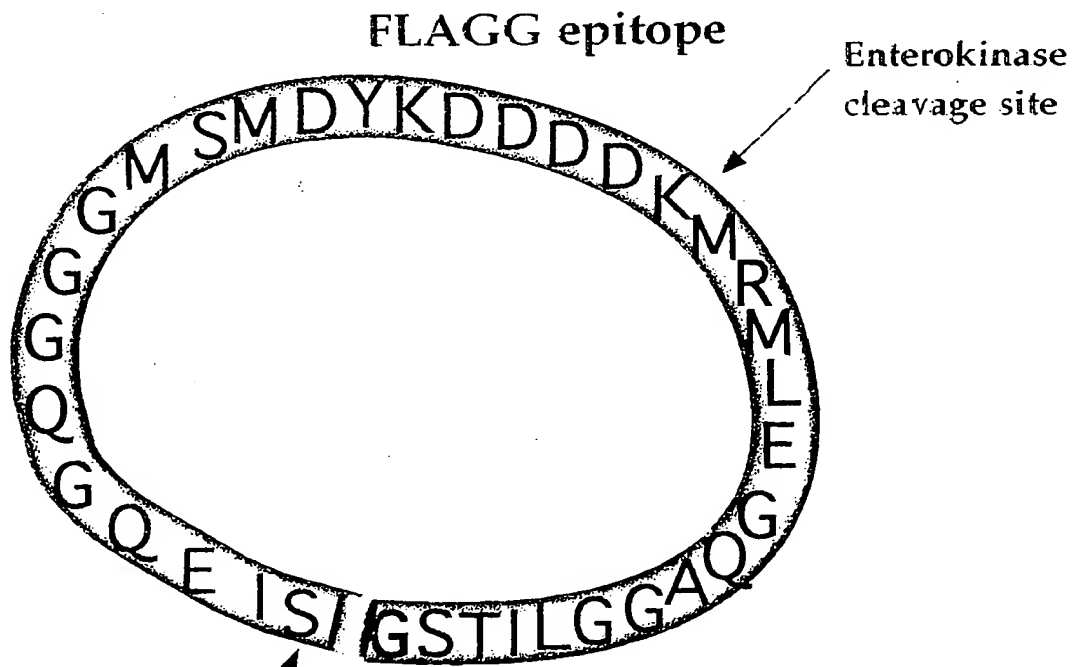
IIVHN[S]IEQGQGGGMSMDYKDDDDKMRMLEGQAGGLITS[G]CIS

GDSLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESKVS RVFCT

## Intein A

GKKLVYILKTRLGRTIKATANHRFLTIDGWKRLDELSLKEHIALPRK

LESSSLQLSIHGYH



This is the only invariant intein-encoded amino acid (depending on intein used this can be a cysteine, serine or threonine).

Figure 4B

CMV promoter →

1/1	31/11	61/21
GCT TCG CGA TGT ACG GGC CAG ATA TAC GCG TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC A		
121/41	151/51	181/61
TAC GGT AAA TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC AAT AAT GAC GTA T		
241/81	271/91	301/101
TTT ACG GTA AAC TGC CCA CTT GGC AGT ACA TCA AGT GTA TCA TAT GCC AAG TAC GCC CCC TAT TGA C		
361/121	391/131	421/141
GGA CTT TCC TAC TTG GCA GTA CAT CTA CGT ATT AGT CAT CGC TAT TAC CAT GGT GAT GCG GTT TTG G		
481/161	511/171	541/181
CCA CCC CAT TGA CGT CAA TGG GAG TTT GTT TTG GCA CCA AAA TCA ACG GGA CTT TCC AAA ATG TCG T		
601/201	631/211	661/221
CTA TAT AAG CAG AGC TCT CTG GCT AAC TAG AGA ACC CAC TGC TTA CTG GCT TAT CGA AAT TAA TAC G		
721/241	751/251	781/261
CTg tcg acT GGA GGA ACC ATG GAG TCC GGA tca cca gaa ata gaa aag ttg tct cag agt gat att t	intB (lc)	
M E S G S P E I E K L S O S D I Y		
841/281	871/291	901/301
ttg act gtg cca gga cca cat aac ttt gtc gcc aat gac atc att gtc cat aac agt ATC GAA CAA g		
L T V P G P H N F V A N D I I V H N S I E O G		
961/321	991/331	1021/341
ATG ctg gag ggc caa gca ggt gga CTG ATC ACC agt ggc TGC ATC AGT GGA GAT AGt ttg atc agc t		
M L E G Q A G G L I T S G C I S G D S L I S L		
1081/361	1111/371	1141/381
ttt gaa ata tgg gca att aat gaa cag acg atg aag cta gaa tca gct aaa gtt agt cgt gta ttt t		
F E I W A I N E O T M K L E S A K V S R V F C		
1201/401	1231/411	1261/421
aag gca aca gca aat cat aga ttt tta act att gat ggt tgg aaa aga tta gat gag cta tct tta a		
K A T A N H R F L T I D G W K R L D E L S L K		
1321/441	1351/451	1381/461
GAT cca tgg tta cca TGA caa ttg GCG GCC GCT CGA GTC TAG AGG GCC CGC GGT TCG AAG GTA AGC C		
D P W L P *		
1441/481		
ATC ACC ATT GAG TTT AAA CCC GCT GAT		

F09020-020966

FIGURE 5

A)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
GACTCCATCGTTTCTATTACGG  
AGAC  
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC  
AATGACATCATTGTCCATAAC  
AGTG  
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
CCATCGGCGACGGCCCCGTGCT  
GCTG  
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
AAGCGCGATCACATGGTCCTGC  
TGGA  
GTTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC  
GAACGGGGAATTCTCGCAGGTA  
GACA  
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG  
TCGAGCTGGACGGCGACGTA  
CGGC  
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
ACCCTGAAGTTCATCTGCACCA  
CCGG  
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG  
TGCTTCAGCCGCTACCCCGAC  
CACA  
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC  
GCACCATCTTCTTCAAGGACGA  
CGGC  
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC  
ATCGAGCTGAAGGGCATCGACT  
TCAA  
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA  
CGTCTATATCATGGCCGACAAG  
CAGA  
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC  
AAAAGCTGATATGCATCTCCGG  
AaAT  
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG  
ATGAAAAAGATTTTGAAATAT  
GGGC  
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT  
ACTGGCAAAAAGCTAGTTTAT  
ATTT  
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA  
CTATTGATGGTTGGAAAAGATT  
AGAT





[illegible]

## 5.4



GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT  
TACAATTAGGCCTCCGCGGCC  
AGTA  
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

**F)**  
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS  
VQLADHYQQNTPIGDG  
PVLL  
PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM  
VSKGEELFTGVVPILVELDG  
DVNG  
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ  
HDFFKSAMPEGYVQERTIFF  
KDDG  
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI  
KVNFKIRHNIEDLEQKLIK  
ISGN  
SLISLASTGKRVS IKDLLDEKDFEIWAVNEQTMKLESAKVS RVFCTGKKLVYILKTRL  
GRTIKATANHRFLTIDGW  
KRLD  
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

**G)**  
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
GACTCCATCGTTTCTATTACGG  
AGAC  
TGGAGTCGAAGAGGTTTTTGATTGgCcGTGCCAGGGCCCCATAACTTTGTGGCCA  
ATGACATCATTGTCCATAAC  
AGTG  
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
CCATCGGCGACGGCCCCGTGCT  
GCTG  
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
AAGCGCGATCACATGGTCCTGC  
TGGA  
GTTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC  
GAACGGGGAATTCTCGCAGGTA  
GACA  
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG  
TCGAGCTGGACGGCGACGTAAA  
CGGC  
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
ACCCTGAAGTTCATCTGCACCA  
CCGG  
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGACG

TGCTTCAGCCGCTACCCCGAC  
 CACA  
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC  
 GCACCATCTTCTTCAAGGACGA  
 CGGC  
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC  
 ATCGAGCTGAAGGGCATCGACT  
 TCAA  
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA  
 CGTCTATATCATGGCCGACAAG  
 CAGA  
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC  
 AAAAGCTGATATGCATCTCCGG  
 AAAT  
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG  
 ATGAAAAAGATTTTGAAATAT  
 GGGC  
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTGT  
 ACTGGCAAAAAGCTAGTTTAT  
 ATTT  
 TAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA  
 CTATTGATGGTTGGAAAAGATT  
 AGAT  
 GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT  
 TACAATTAGGCCTCCGCGGCC  
 AGTA  
 CCCCTACGACGTCCCGGACTACGCTATCGATTAA

**H)**  
 MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLAVPGPHNFVANDIIVHNSEEDLGSS  
 VQLADHYQQNTPIGDG  
 PVLL  
 PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM  
 VSKGEELFTGVVPILVELDG  
 DVNG  
 HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ  
 HDFFKSAMPEGYVQERTIFF  
 KDDG  
 NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGI  
 KVNFKIRHNIEDLEQKLIC  
 ISGN  
 SLISLASTGKRVSIGDLLDEKDFEIWAINETMKLESASVSRVFCTGKKLVYILKTRL  
 GRTIKATANHRFLTIDGW  
 KRLD  
 ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

D)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
 GACTCCATCGTTCTATTACGG  
 AGAC  
 TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC  
 AATGACATCATTGTCCATAAC  
 AGTG  
 AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
 CCATCGGCGACGGCCCCGTGCT  
 GCTG  
 CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
 AAGCGCGATCACATGGTCCTGC  
 TGGA  
 GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTG  
 GAACGGGGAATTCTCGCAGGTA  
 GACA  
 AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG  
 TCGAGCTGGACGGCGACGTAAA  
 CGGC  
 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
 ACCCTGAAGTTCATCTGCACCA  
 CCGG  
 CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG  
 TGCTTCAGCCGCTACCCCGAC  
 CACA  
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC  
 GCACCATCTTCTTCAAGGACGA  
 CGGC  
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC  
 ATCGAGCTGAAGGGCATCGACT  
 TCAA  
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAA  
 CGTCTATATCATGGCCGACAAG  
 CAGA  
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC  
 AAAAGCTGATATGCATCTCCGG  
 AaAT  
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG  
 ATGAAAAAGATTTTGAAATAT  
 GGGC  
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT  
 ACTGGCAAAAAGCTAGTTTAT  
 ATTT  
 TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA  
 CTATTGATGGTTGGAAAAGATT  
 AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT  
TACAATTAGGCCTCCGCGGCC  
AGTA  
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

**J)**  
MESGSPEIEKLSQSDIYWDSIVPITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS  
VQLADHYQQNTPIGDG  
PVLL  
PDNYHLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM  
VSKGEELFTGVVPILVELDG  
DVNG  
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ  
HDFFKSAMPEGYVQERTIFF  
KDDG  
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI  
KVNFKIRHNIEDLEQKLIK  
ISGN  
SLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESAKVS RVFCTGKKLVYILKTRL  
GRTIKATANHRFLTIDGW  
KRLD  
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

**K)**  
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
GACTCCATCGTTTCTATTACGG  
AGAC  
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC  
AATGACATCATTGTCCATAAC  
AGTG  
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
CCATCGGCGACGGCCCCGTGCT  
GCTG  
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
AAGCGCGATCACATGGTCCTGC  
TGGA  
GTTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC  
GAACGGGGAATTCTCGCAGGTA  
GACA  
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG  
TCGAGCTGGACGGCGACGTAAA  
CGGC  
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
ACCCTGAAGTTCATCTGCACCA  
CCGG  
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

TGCTTCAGCCGCTACCCCGAC  
 CACA  
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC  
 GCACCATCTTCTTCAAGGACGA  
 CGGC  
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC  
 ATCGAGCTGAAGGGCATCGACT  
 TCAA  
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA  
 CGTCTATATCATGGCCGACAAG  
 CAGA  
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC  
 AAAAGCTGATATGCATCTCCGG  
 AaAT  
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG  
 ATGAAAAAGATTTTGAAATAT  
 GGGC  
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT  
 ACTGGCAAAAAGCTAGTTTAT  
 ATTT  
 TAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA  
 CTATTGATGGTTGGAAAAGATT  
 AGAT  
 GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAACTAGAAAGCTCCTCTT  
 TACAATTAGGCCTCCGCGGCC  
 AGTA  
 CCCCTACGACGTCCCGGACTACGCTATCGATTAA

L)  
 MESGSPEIEKLSQSDIYWDSIVSITETGVVEEVDLTVPGPHNFVANDIIVHNSEEDLGSS  
 VQLADHYQQNTPIGDG  
 PVLL  
 PDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM  
 VSKGEELFTGVVPILVELDG  
 DVNG  
 HKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKQ  
 H DFFKSAMPEGYVQERTIFF  
 KDDG  
 NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGI  
 KVNFKIRHNIEDLEQKLIC  
 ISGN  
 SLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESAKVS RVFCTGKRLVYILKTRL  
 GRTIKATANHRFLTIDGW  
 KRLD  
 ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

M)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
 GACTCCATCGTTTCTATTACGG  
 AGAC  
 TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC  
 AATGACATCATTGTCCATAAC  
 AGTG  
 AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
 CCATCGGCGACGGCCCCGTGCT  
 GCTG  
 CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
 AAGCGCGATCACATGGTCCTGC  
 TGGA  
 GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC  
 GAACGGGGAATTCTCGCAGGTA  
 GACA  
 AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG  
 TCGAGCTGGACGGCGACGTAAA  
 CGGC  
 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
 ACCCTGAAGTTCATCTGCACCA  
 CCGG  
 CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGAG  
 TGCTTCAGCCGCTACCCCGAC  
 CACA  
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC  
 GCACCATCTTCTTCAAGGACGA  
 CGGC  
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGC  
 ATCGAGCTGAAGGGCATCGACT  
 TCAA  
 GGAGGACGGCAACATCCTGGGGGCACAAGCTGGAGTACAACAGCCACAA  
 CGTCTATATCATGGCCGACAAG  
 CAGA  
 AGAACGGCATCAAGGTGAAGTTCAAGATCCGCCACAACATCGAGGACCTCGAGC  
 AAAAGCTGATATGCATCTCCGG  
 AGAT  
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG  
 ATGAAAAAGATTTTGAAATAT  
 GGGC  
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT  
 ACTGGCAAAAAGCTAGTTTAT  
 ATTT  
 TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAaATTTTAAAC  
 TATTGATGGTTGGAAAAGATT  
 AGAT •

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT  
TACAATTAGGCCTCCGCGGCC  
AGTA  
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

N)  
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS  
VQLADHYQQNTPIGDG  
PVLL  
PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM  
VSKGEELFTGVVPILVELDG  
DVNG  
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ  
HDFFKSAMPEGYVQERTIFF  
KDDG  
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI  
KVNFKIRHNIEDLEQKLIK  
ISGD  
SLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESAKVS RVFCTGKKLVYILKTRL  
GRTIKATANHKFLTIDGW  
KRLD  
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

O)  
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
GACTCCATCGTTCTATTACGG  
AGAC  
TGGAGTCGAAGAGGTTTTTGATTGACTGTGCCAGGGCCCCATAACTTTGTGGCC  
AATGACATCATTGTCCATAAC  
AGTG  
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
CCATCGGCGACGGCCCCGTGCT  
GCTG  
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
AAGCGCGATCACATGGTCCTGC  
TGGA  
GTTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC  
GAACGGGGAATTCTCGCAGGTA  
GACA  
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG  
TCGAGCTGGACGGCGACGTAAA  
CGGC  
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
ACCCTGAAGTTCATCTGCACCA  
CCGG  
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGACG





Q)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG  
 GACTCCATCGTTTCTATTACGG  
 AGAC  
 TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC  
 AATGACATCATTGTCCATAAC  
 AGTG  
 AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC  
 CCATCGGCGACGGCCCCGTGCT  
 GCTG  
 CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG  
 AAGCGGATCACATGGTCCTGC  
 TGGA  
 GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC  
 GAACGGGGAATTCTCGCAGGTA  
 GACA  
 AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG  
 TCGAGCTGGACGGCGACGTAAA  
 CGGC  
 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG  
 ACCCTGAAGTTCATCTGCACCA  
 CCGG  
 CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG  
 TGCTTCAGCCGCTACCCCGAC  
 CACA  
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC  
 GCACCATCTTCTTCAAGGACGA  
 CGGC  
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGC  
 ATCGAGCTGAAGGGCATCGACT  
 TCAA  
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAA  
 CGTCTATATCATGGCCGACAAG  
 CAGA  
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC  
 AAAAGCTGATATGCATCTCCGG  
 AGAT  
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG  
 ATGAAAAAGATTTTGAAATAT  
 GGGC  
 AGTTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT  
 ACTGGCAAAAAGCTAGTTTAT  
 ATTT  
 TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA  
 CTATTGATGGTTGGAAAAGATT  
 AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT  
TACAATTAGGCCTCCGCGGCC  
AGTA  
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

**R)**  
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS  
VQLADHYQQNTPIGDG  
PVLL  
PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM  
VSKGEELFTGVVPILVELDG  
DVNG  
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ  
HDFFKSAMPEGYVQERTIFF  
KDDG  
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI  
KVNFKIRHNIEDLEQKLIC  
ISGD  
SLISLASTGKRVS IKDLLDEKDFEIWAVNEQTMKLESAKVS RVFCTGKKLVYILKTRL  
GRTIKATANHRFLTIDGW  
KRLD  
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

TOP SECRET 02203869

CMV promoter 

**TAP**

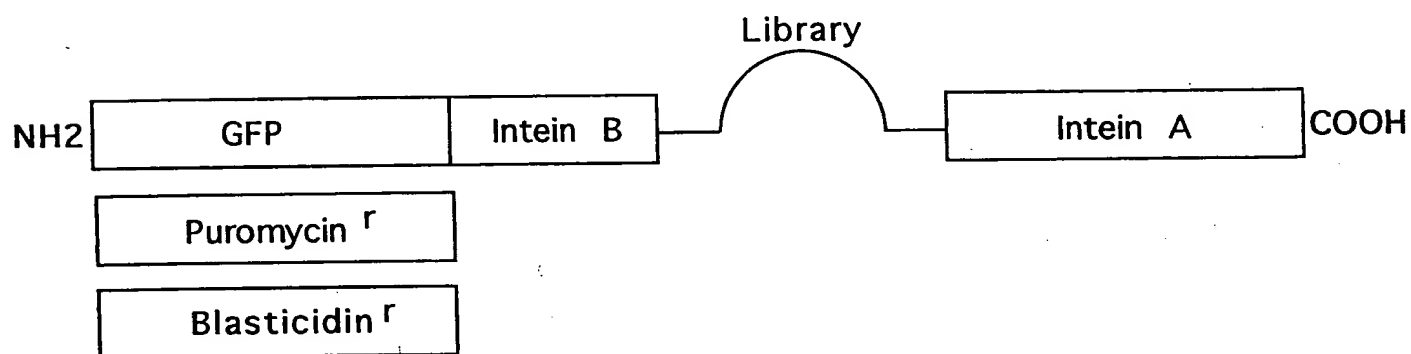
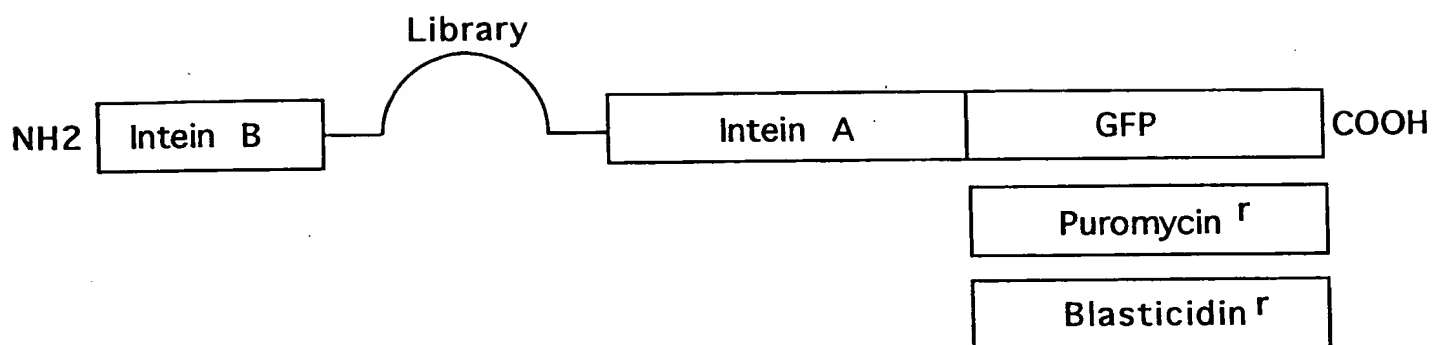


Figure 7



Membrane-tethered scaffold

1. Use a PCR mutagenesis or shuffling approach to mutate intein domains
2. Create a retroviral library of mutants
3. Infect cells and screen for those most efficient at cyclization (assayed indirectly by monitoring the release of IntA-TetRVP16 from its membrane location)

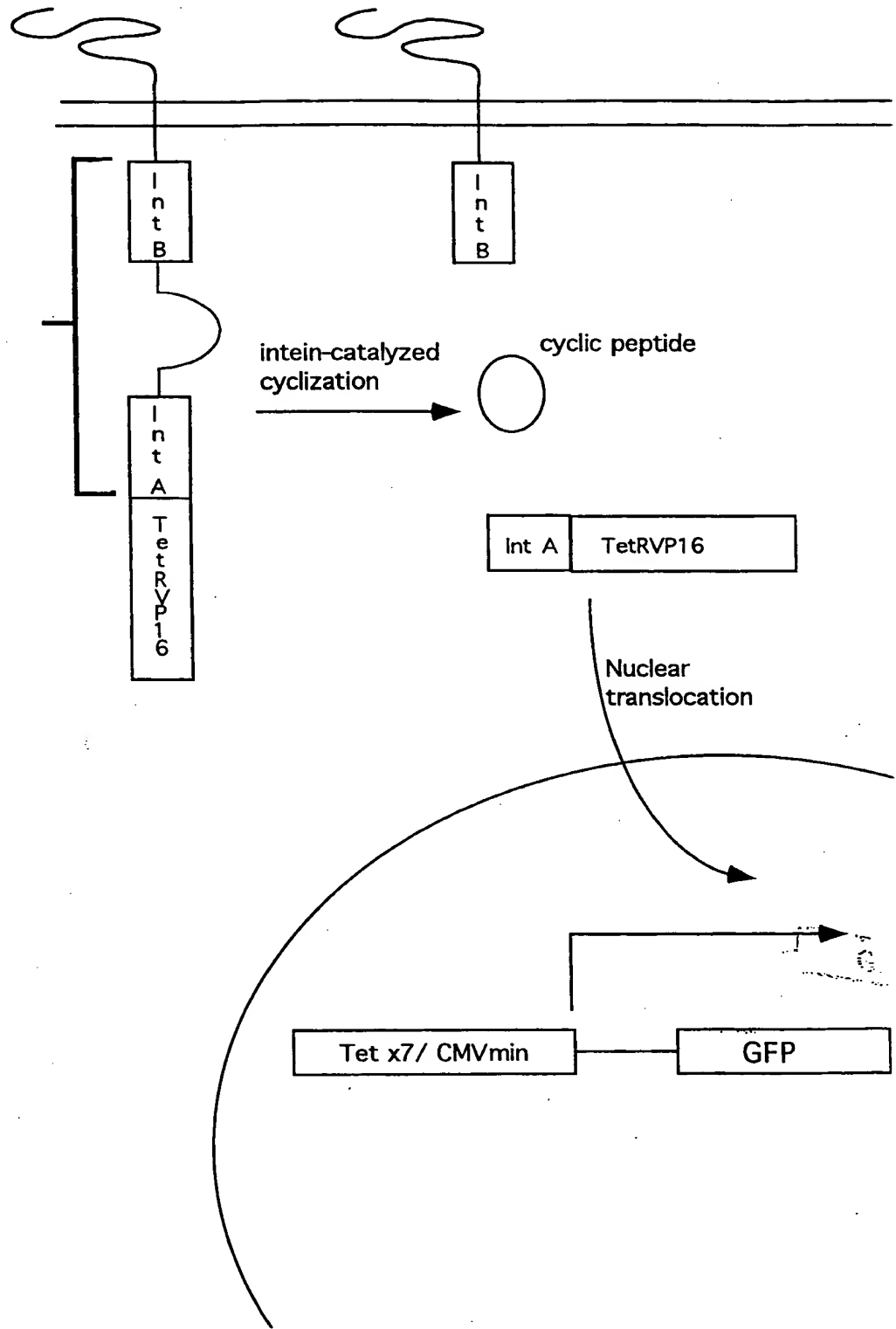
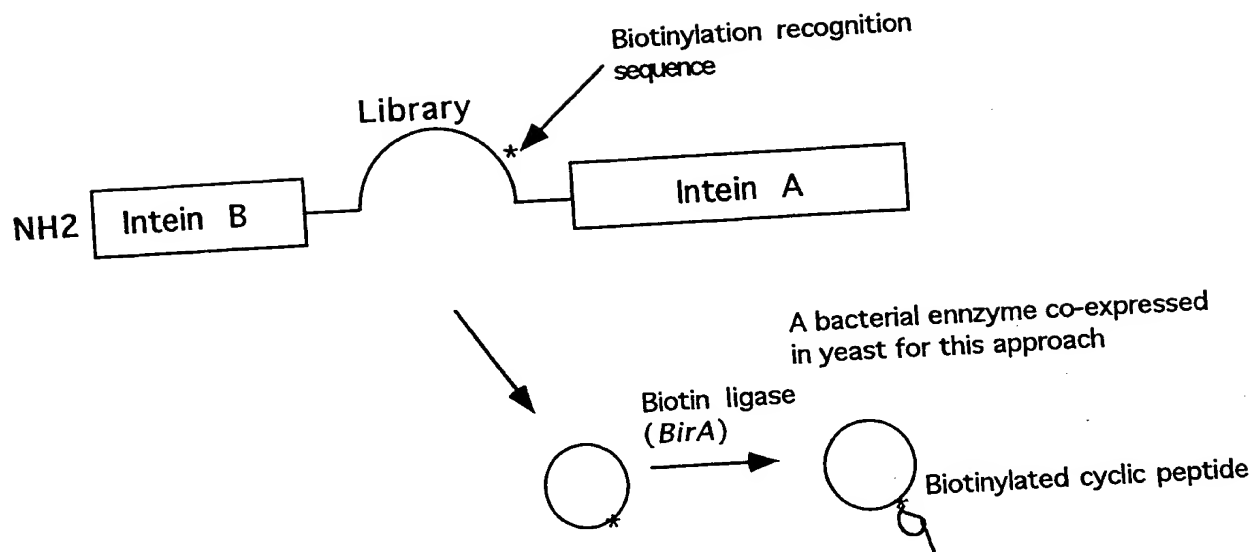


Figure 9



typical cDNA target/transactivation domain fusion utilized in yeast two hybrid systems

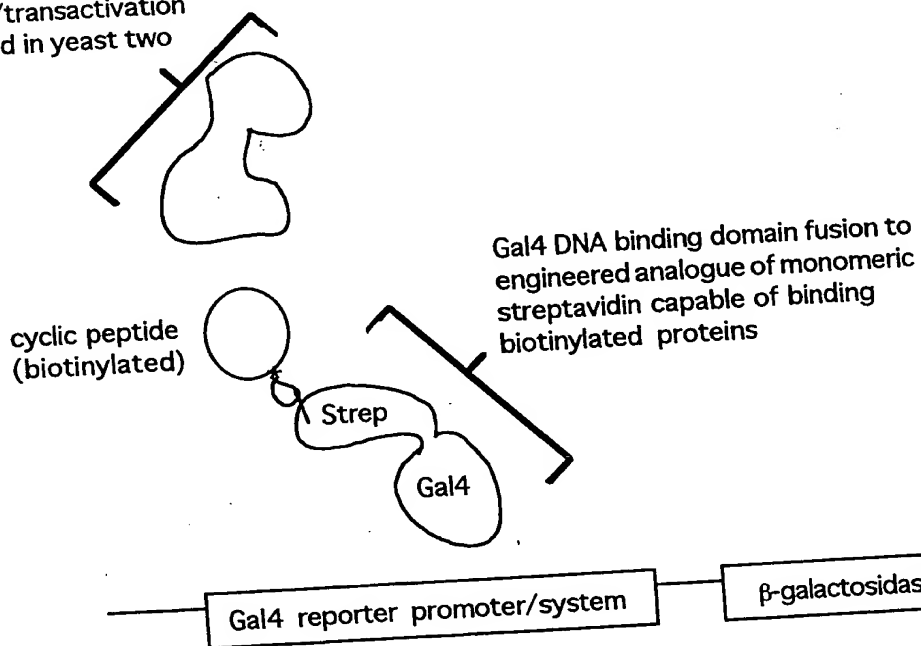
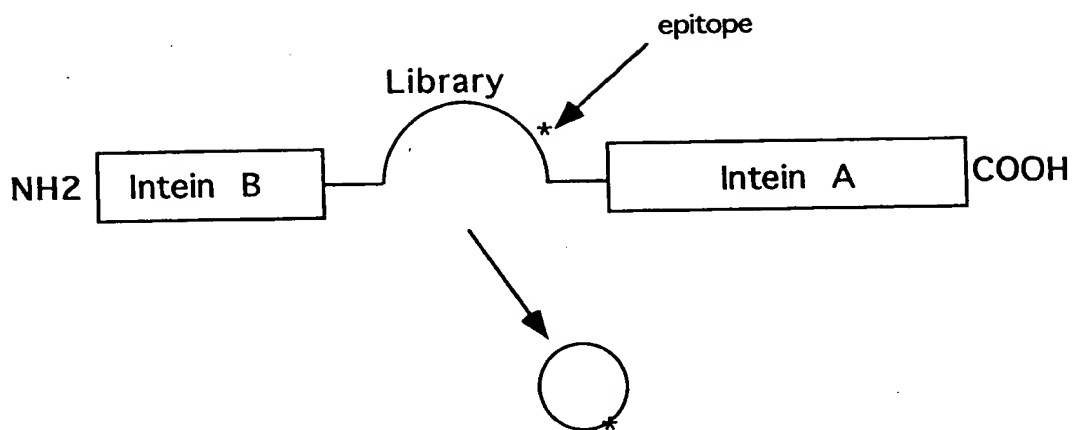
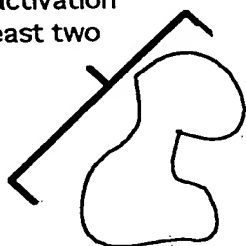


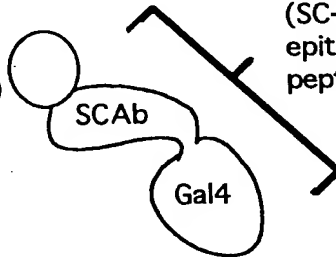
Figure 10



typical cDNA target/transactivation domain fusion utilized in yeast two hybrid systems



cyclic peptide (with epitope)



Gal4 DNA binding domain fusion to engineered single chain antibody (SC-Ab) capable of binding to the epitope present within the cyclized peptide

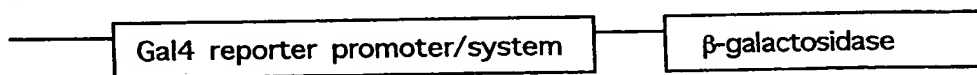
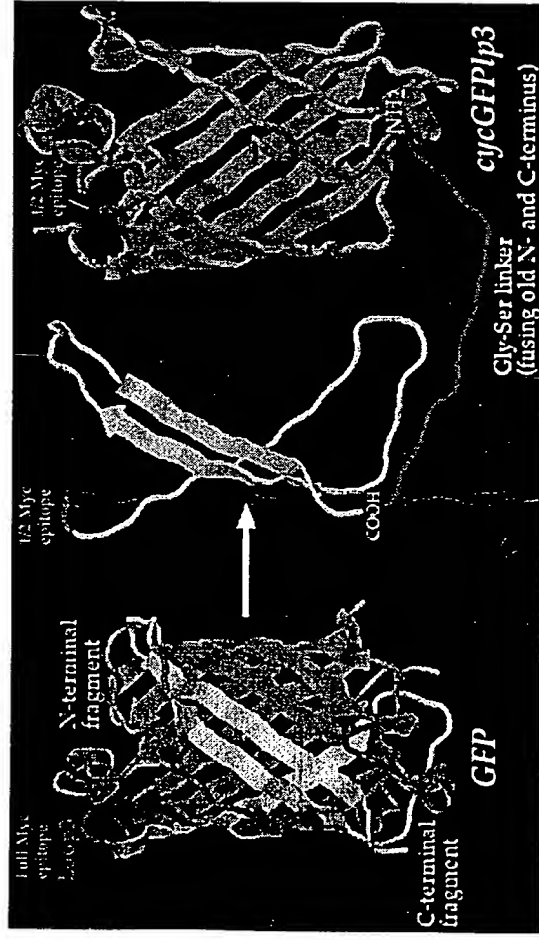


Figure 11



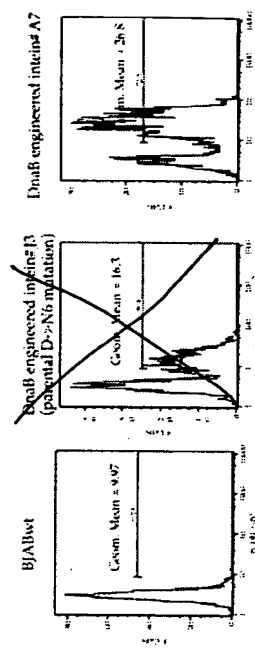
A.



B.

Int B (1c)  
 HSEGEFEKLSQSDLYVDSDIVSITETGVEEVFDLTPVGP  
 HNFVANDYVHSEEDGSSVOLADHYQONTPIGDPVLL  
 PDNHYLSTQSALS KDPNEKRDHNVLEFVTAAGITLGMDE  
 LYKGSNGEFSQVPSKSHVSKGEELFTGVVPLVLDGDVNG  
 HKFSVSGEGEGDATYGLTLKLFICTTCKLPVWPFTLVTTL  
 TYGVOCFSRYPDHMKOHDFPKSAMPEGVVOERTIFFKDDG  
 NYKTRAEVKFEGDTLVNRILLKGIKEDKGNILGHKLEYE  
 YNSHNVYTHADKOKNGIKVNFKIRHNIED<sup>MYC</sup> EQRHICISCD  
 SLTSLASTGKRVSKDLDDEDFEINAINETMKESAKV  
 SRVPTCKKLVYPTCKTRDGRITATANHRELYTIDGKRLD  
 ELSLKEHTALPRKLESSLQGLRGQVPYDVPDYAIG  
 HA

D.



E.

F.

C.

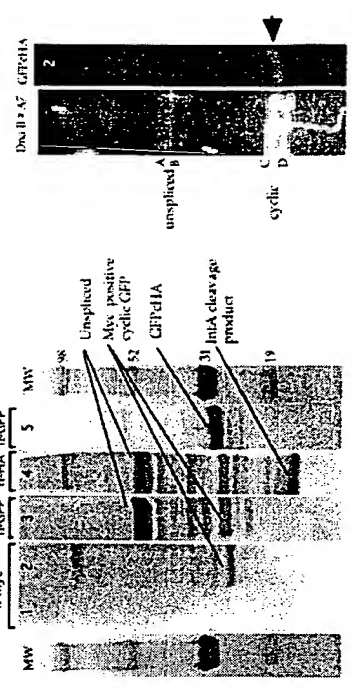
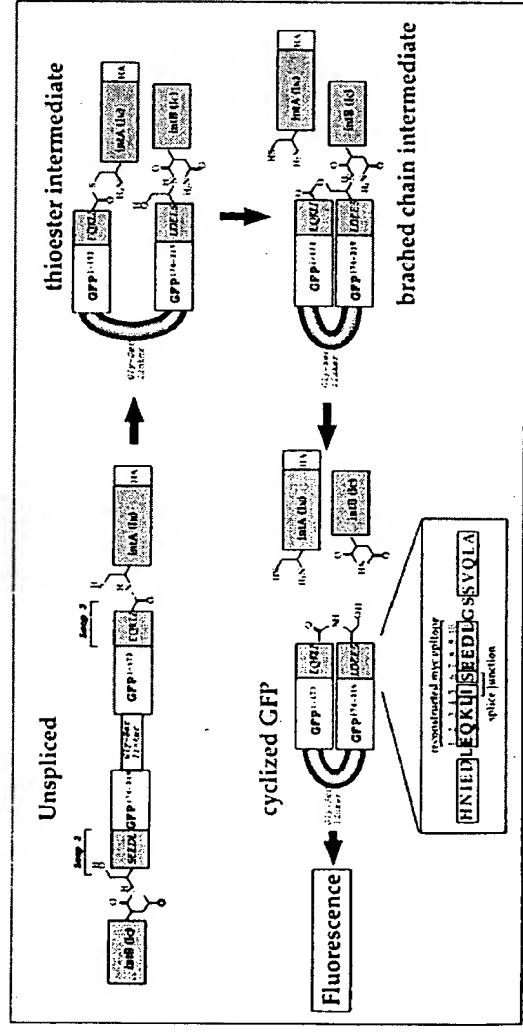
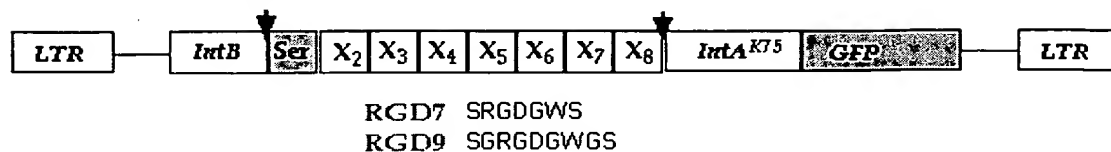


Figure 12



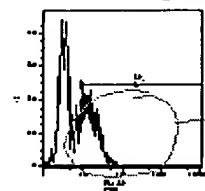


Retroviral library construct: control RGD inserts



Infect cell line

sort GFP<sup>+</sup> population



establish cell line

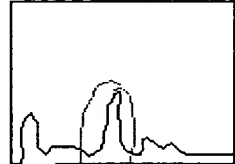
LTR

Lysate

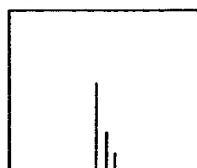
10kDa MWCO

10kDa lysate

collect fraction



MALDI-TOF  
expected mass



C18

RP-HPLC

cyclic  
standard

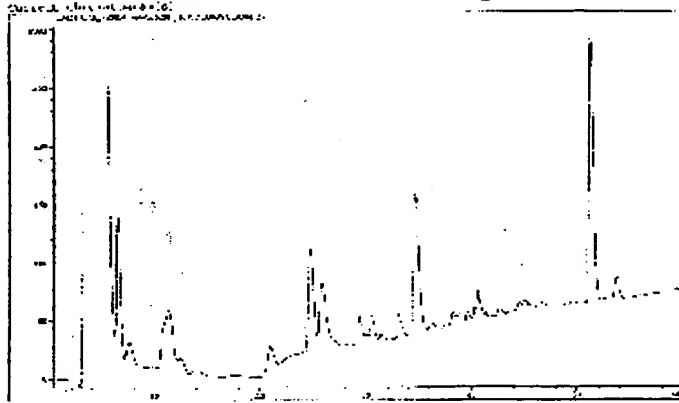
C18

retention  
time  
standard



Figure 15A

# A5T4-RGD7 lysate



A5T4-RGD7 lysate (HPLC 34-35min. fraction standard)

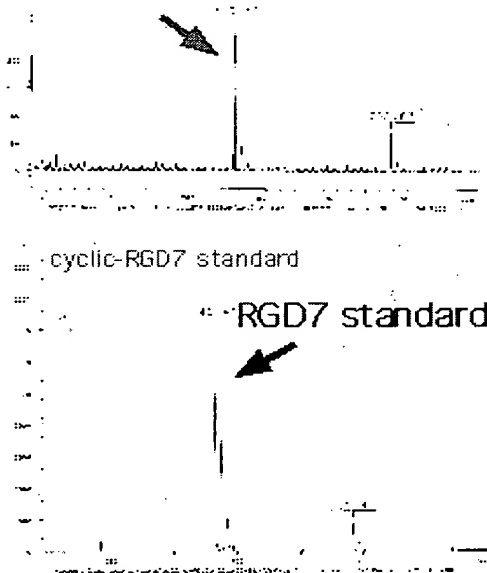
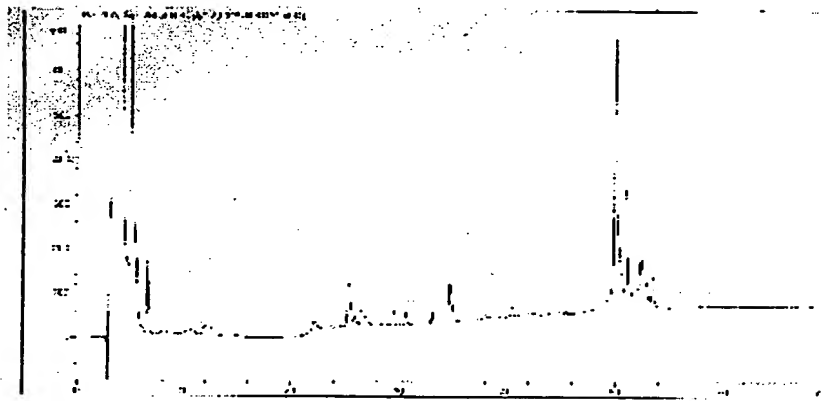


Figure 15 B

# A5T4-RGD9 lysate



A5T4-RGD9 lysate (HPLC 33-34 min.  
fraction standard (expect: 860.4)

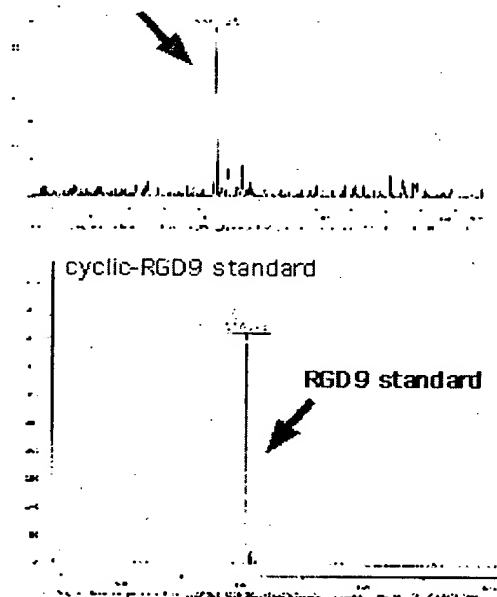


Figure 15C

# LC/MS fragmentation fingerprinting

0128/01 05:57:43 PM

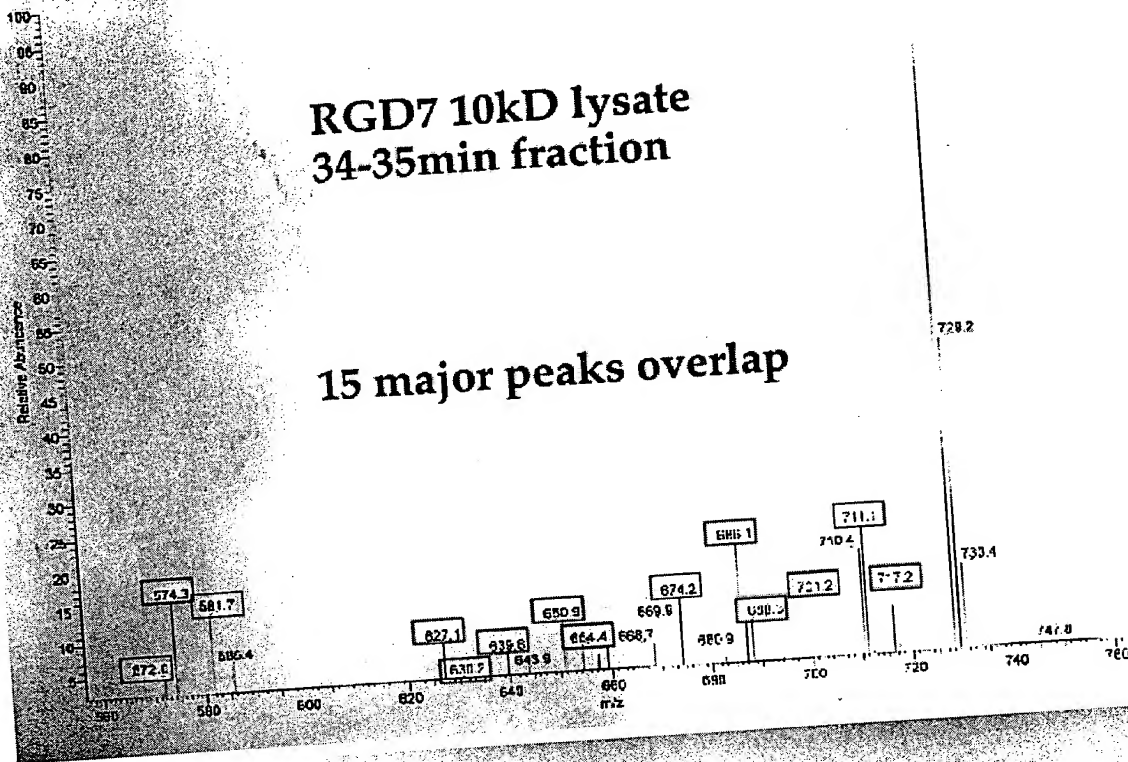
D:\LCQ Data\blank\_star\_RGD7

Blank\_star\_RTD: #1766 RT: 38.50 AM: 1 NL: 4.05E4  
T: + c d Full m/z 748.48 @ 32.00 (185.00-1535.00)

728.2

RGD7 10kD lysate  
34-35min fraction

15 major peaks overlap



D:\LCQ Data\RGD7\_Std

01/28/01 04:35:00 PM

RGD7 std #417 RT: 4.70 AM: 1 NL: 1.50E4  
T: + c d Full m/z 742.70 @ 32.00 (185.00-1535.00)

RGD7 standard

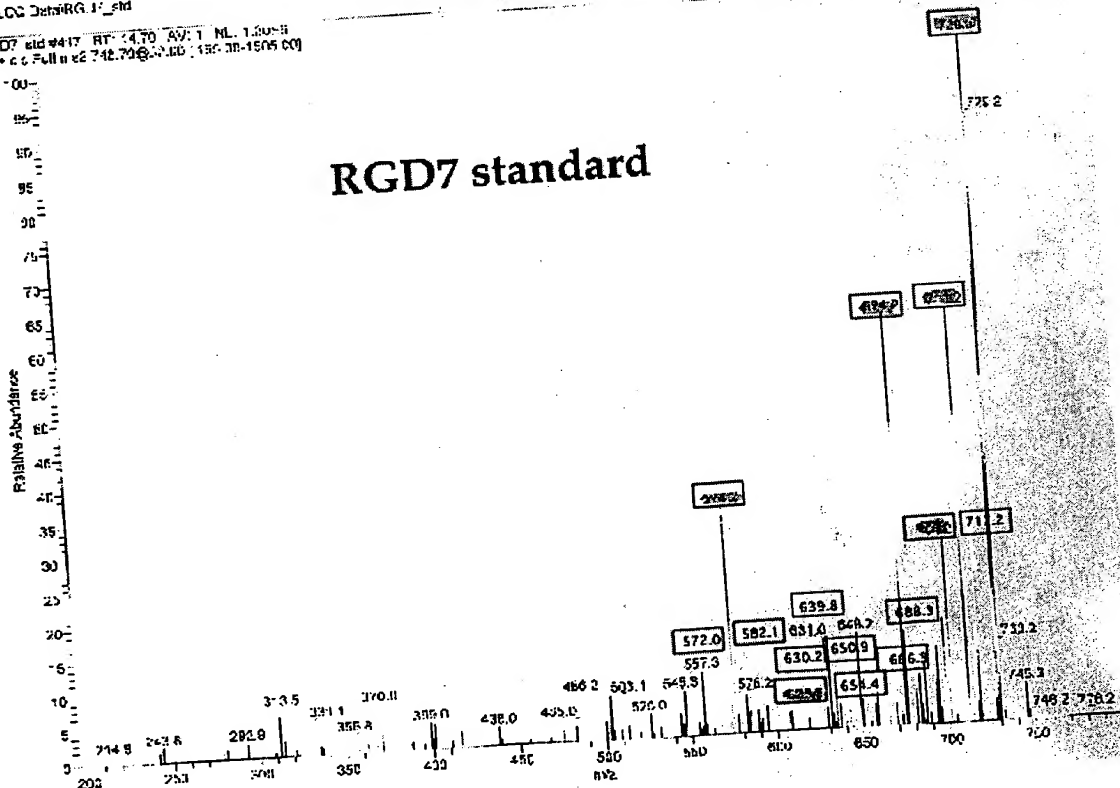


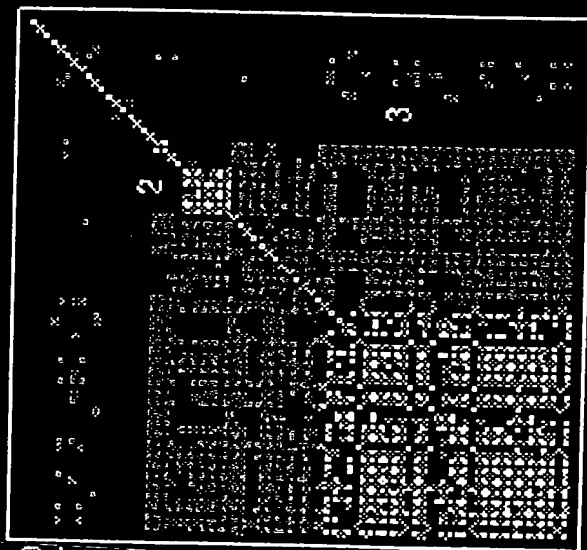
Figure 15D





UNITED STATES GOVERNMENT PRINTING OFFICE: 1977 O - 340-000

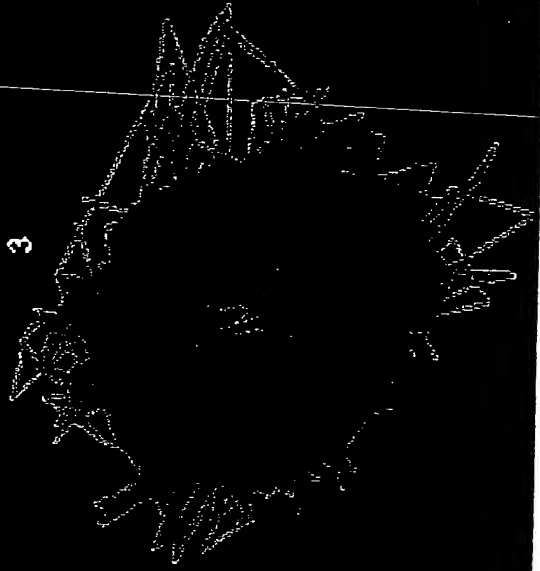
cyclic[SRGPGWS]



11.0 77.0

Frame

RMSD (Å)  
 0.00-1.00  
 1.00-2.00  
 2.00-3.00  
 3.00-4.00



2



Figure 17